CSML IFREMER Brest workshop
6-7 march 2008
CERSAT meeting room
Minutes

Table of Contents

1 Why a CSML workshop at IFREMER ?

During the EuroDESS/EuroMISS/DMS ECOOP meeting at HR Wallingford on 21-22 January 2008, the idea of organising a workshop between CSML designers and Marine Data Management System designers was agreed between HR Wallingford (which is close to CSML team), CMRC and IFREMER.

The Marine Data Management System designers are interested in knowing more about CSML and are expecting of it a better harmonisation/integration of the common features provided in operational oceanography (grids, profiles, sections, ...).

Likewise, the CSML design team are interested in a better understanding of the data management requirements for operational oceanography.

2 What is CSML ?

(A short introduction by Andew Woolf)
CSML is a research output of the NERC DataGrid project.

The research work on the modelling language started in 2003. It was then based on ISO standards for time, location, ... modelisation.

In 2004, based on BODC and BADC requirements for data management, 36 feature types were defined. These were eventually reduced to just 7 feature types.

CSML version 1 was released in 2005. It was based on the concepts of :

- domain reference : which is the origin location of the observation (e.g. the sensor).
- domain complement : which is an offset away from the location where the sensor actually measure properties of the feature (e.g. ocean or atmosphere).

However because this model was weakly typed, it was not useful for developing tools. Most of the observation were 'ProfileFeatures' (radar, CTD, ...). This split geo-spatial domain also proved difficult to work with.

The "Exeter Communiqué" established the idea of 'governance' as being a deterministic factor in the definition of feature types, along with the idea of affordances (https://www.seegrid.csiro.au/twiki/pub/Marineweb/MOTIIVE/OGC_Web_Services_and_GML_Modelling_for_Operational_Meteorology--communique_V1.pdf).

In 2007, the CSML version 2 was released. This version was more strongly typed than version 1 with 13 feature types to enable better distinction between different profiles for example. This version is also compliant with the O&M pattern. With this version, the storage description is separated from the GML encoding.

CSML is now coming out of research activity. Version 3 will be released by the end of this year.

The CSML design team is open to the requirements of the ocean/meteo community and the information model may be tuned for that purpose.

3 European and international framework, INSPIRE

(presentation by Andew Woolf)

INSPIRE is a European directive which contains :

- Processes and procedures for data management system (design, operation)
- Requirements for metadata, spatial data sets, network services
- Requirements for data service, monitoring.

Next year, the Ocean and Atmosphere models integration within INSPIRE will be studied by an official working group.
Anyone can propose candidates, but the number of them is limited (in order to improve efficiency).

Because the Ocean and Atmosphere models will be specified by INSPIRE, the lifetime of CSML is expected to be another 2 years. However, the CSML designers plan to disseminate CSML principles and concepts to the INSPIRE Ocean and Atmosphere working groups. Thus, CSML will have an influence on prospective INSPIRE Ocean and Atmosphere models.

In terms of INSPIRE data specification compliance, internal data flows within an organisation do not need to be compliant, but the data flow to external users must be compliant.

4 Presentation of projects

The slides used for presentation of different projects are available on ftp :

ftp://ftp.ifremer.fr/ifremer/sismer/csml/brest_meeting_20080306/projects

project/ocean data management systems	**Who ?**	**file**
CORIOLIS	*T. Carval*	cordo-tra-08-033-presentation-csml-coriolis.ppt
PREVIMER	*T. Carval*	cordo-tra-08-033-presentation-csml-previmer.ppt
CERSAT	*J.F. Piollé*	CERSAT_20080306__CSML_Brest.ppt
SEADATANET	*V. Harscoat*	csml_meeting_vh_seadatanet_060308.ppt
MERSEA/MYOCEAN	*T. Loubrieu*	CSML-MyOceanDataDescription.ppt
ECOOP	*D. Dunne*	ECOOP_CSML_IFREMER.ppt
INTERRISK	*D. Dunne*	InterRisk_CSML_IFREMER.ppt
ATOLL	*O. Lauret*	Presentation_20080307-ATOLL-P2-UK2.ppt
SEXTANT/Seagrid	*M. Treguer*	Seagrid.ppt

Some specific project requirements are :

- wave spectra management for CERSAT, PREVIMER (at one location, energy measurement with 2 dimension : direction and wavelength).
- management of different level of processing from sensor signal to geophysical parameters for ATOLL.

- possibility of requesting a dataset by parts (split request results and get subset 1st/3, then 2nd/3, at last 3rd/3) (for MERSEA/MyOCEAN).
- the same observation may be disseminated with different data model (more specific, to more generic), depending on the usage of the data and the compilation of of heterogenous observation within one product (e.g. for numerical data model assimilation) (for CORIOLIS).

- The SEXTANT/Seagrid server tool has been presented. This OGC server enables dissemination of netCDF/CF gridded dataset in WCS/WMS. A specific interpretation of the WMS protocol has been used for managing colormap (specific extra-dimension called 'dim_range'). This would have to be harmonised with 4D (or grids series) objects data management.

An overview of the data management projects is described in a MS Excel file available on ftp: ftp://ftp.ifremer.fr/ifremer/sismer/csml/brest_meeting_20080306/projects/project_breakdown.xls

5 CSML detailed presentation/demonstration

(presentation by Andew Woolf, demonstration by Dominic Lowe)

CSML is a modelling language which is supported by tools for serialising particular data formats (e.g. netCDF, RDBMS, ...) into GML.

The projects / standards which are close to CSML (partly overlapping, partly complementary) are : the Common Data Model (of UNIDATA) and O&M standard (of OGC).

MarineXML is an 'umbrella' for different projects/model which deals with ocean data modeling. CSML participated in MarineXML.

CSML focuses on geo-physical/geo-chemicals features (because it is not possible to cover everything).

The key object in the model is the 'featureType' which is the generic object for describing specific featureType and hosting climate science feature. The featureType is really like an 'object' (in object modelling). The featureType properties are :

- attributes
- operations
- associations (with others objects)

FeatureType are distinguished on the basis of the geometry of the observation.

CSML defines 13 featureTypes in V2 (Point, PointSeries, Trajectory, PointCollection ...).

A CSML feature (of type 'FeatureType') describes a phenomenon (parameter) on a domain (spatio temporal coordinates) with a value range (which is the actual value of the parameter on the domain).

The CSML coverage is the mapping from some spatio-temporal domain to a value range.

Most of the vocabulary used to define the CSML concepts are taken out of the GML concepts.

The governance is the main driver for granularity. If some existing authority is responsible for a delimited field, then a data model or a data model specification may describe that field. Governance is very important. For example :

- CSML deals with geo-physical ocean and atmosphere features.
- Parameters names are described in vocabularies which are define somewhere else (and can be changed depending on the requirements).
- The coordinates systems are mainly taken out EPSG classification.

We have to consider apart of CSML, the process of observation (sensor signal, sensor description...) which is part of the Sensor Web Enablement work (SensorML, SOS, ...). CSML deals with result of observations.

The CSML is overlapping with some part of O&M and very similar to CDM. The teams are already collaborating. The CSML objective for V3 are :

- CSML is fully O&M compliant.
- CSML and CDM scientific data models are merged.

The main way of using CSML (with tools provided by CSML design team) is :

- describe CSML feature with a CSML markup language. This CSML markup language uses extensively the xlink:href attributes to point external datasets in existing data format (netCDF, ...). CSML markup language provides function for data aggregation (which is quite similar to ncml but independent from the data storage).
- use a python library to load data model instances from the CSML file. This python library (or CSML parser) uses plugins which can read the xlink:href in order to load data model instances from the external resources as well. From python, the CSML objects can be used as needed. The actual values of the CSML feature attributes are only read (from external resource for example) if required.
- use python library for serialising data model instances in netCDF files...
- use other python library (with python module for the web) to serialise datasets (feature) in GML within WFS or WMS or netCDF within WCS...

A demonstration of these tools has been made.

6 Data management tools being possible candidates for using the CSML model

6.1 Dap4Cor

(presentation by Thomas Loubrieu)

Dap4cor is an OPeNDAP data server which is used to access vertical profiles from CORIOS database with the OPENDAP/DAPPER in-situ observation convention.

The presentation of Dap4cor is available on ftp :

ftp://ftp.ifremer.fr/ifremer/sismer/csml/brest_meeting_20080306/tools/boulder-dap4cor_presentation.ppt

Dap4cor is a web application (tomcat application) which translates OPeNDAP requests so that a vertical profiles objects instance can be populated from CORIOLIS Oracle RDBMS and serialised in OPeNDAP data format.

Dap4cor use an 'ocean business' layer which only deals with vertical profiles. For the V2 of the server, the planned updates are :

- deals with other ocean feature types.
- enable OGC compliant protocols (WFS).

CSML maybe a good candidate for doing that.

Comparing to the existing 'profile' model in Dap4cor and the RaggedProfileSeries feature type in CSML are very similar but no technical feature (dealing with the management of data, however the data store is) must appear into this data model (which is more or less what happens in Dap4cor).

Note: The CSML/WFS criteria on request does not work as a specific request send to a whole dataset (which subset itself), but as a filter on the whole dataset.

to be careful : There must be a different models for the software implementation which manages the data model (i.e. Dap4cor design) and the data model itself (i.e. CSML).

6.2 ATOLL

(presentation by Olivier Lauret)

Is the overall system for managing datasets which are managed at CLS for the AVISO altimetry data centre.

This deals with metadata of every datasets and files stored at CLS and enable the request/download (through OPeNDAP and netCDF file generation) of subsets.

CSML may be used to describe the datasets managed within that tool. However one part of the datasets are raw data (level of processing below 1B ???). CSML is not designed for managing such dataset but SWE standard aims at doing so.

6.3 ECOOP

(presentation by Declan Dunne)

The ECOOP project is considering implementing WFS (Web Feature Service) for oceanographic in situ data. The CSML data model may be a good candidate for GML Application Schema implementation.

6.4 InterRisk

(presentation by Declan Dunne)

The InterRisk project has implemented a first version of GML Application Schemas for oil spill and harmful algal bloom (HAB) data types to be delivered via WFS (Web Feature Service). These schemas are based on the Arc Marine data model. The CSML data model may also be a good candidate for GML Application Schema implementation. A review of these first version schemas will be carried out by InterRisk in the context of CSML.

7 The Interface principle

The good way of implementing CSML data model within a data server such as DAP4Cor would be the usage of an agreed java interface (based on a data model like CSML). The CORIOLIS data centre may then implements these interfaces so that this data model can be populated (online) from the CORIOLIS Oracle Database. Then the model may be serialised in GML (by using CSML common tools) or in other protocols such as OPeNDAP (by doing developments which then may be used in the ocean/atmosphere community).

No CSML java interface is not yet available but could be a good technical mean for managing the merging/compliancy of the CDM data model and tools with CSML data model and tools. What could be proposed for CSML V3 is a data model compliant with the Common Data Model and shared java interfaces for defining these compliant models.

Being developed mainly in java, the UNIDATA tools for implementation of the Common Data Model could rely this shared and agreed java interface. The compliancy between ocean/atmosphere data management tools (from CSML and UNIDATA) may be achieved on the base of an agreement on the java interface.

This java interfaces may be a way of integrating ocean/atmosphere data store in GeoServer. Then each implementation of this ocean/atmosphere interface could be considered as a ocean/atmosphere data driver dedicated to specific datastorage (netCDF, RDBMS with specific table schema, ODV, ...).

The CSML team may provide for that purpose the java interfaces and an implementation for reading CSML markup language and serialising in GML. Any other implementation may be written depending on data providers requirements, for example :

- Jon Blower from University if Reading could implement the GridsSeries featureType for netCDF files.
- Thomas Loubrieu from IFREMER/CORIOLIS will implement the in-situ observation relevant featureType for CORIOLIS RDBMS.
- ...

8 Roadmap

(animated by Keiran Millard)

8.1 Agreements

Everyone would like a GML AS for 'scientific data types'.
CSML is a 'ready to go' option that has been developed to be compliant with OGC and ISO TC 211 and has been aligned with INSPIRE (Motiive project)
There is ongoing collaboration/engagement between CSML and the Galeon community to examine alignments with NetCDF
CSML would benefit from wider tooling / clients to widen its adoption by users
ECOOP needs to adopt some pragmatic solutions in support of a bigger picture

8.2 ECOOP Related

The use of CSML in ECOOP has been discussed during the meeting.
Main point is : Data from Seprise (initially through HRW, and later via the main ECOOP DMS) and CORIOLIS are going to provide a WFS/CSML output to their datasets, which is expected to be implemented in GeoServer configured with 'complex features'.

8.3 Bigger Picture (MyOcean and beyond - target 12/08)

- CSML V3 (12/08)
  - Investigate "Interface Principle" between CSML / CDM / O&M - STFC (04/08)
  - Definition of interface classes for the above - STFC / UREADESS / Ifremer
    - Java implementation (?)
    - Python implementation (?)
- Geoserver V1.6l
  - "Tidy" O&M implementation now O&M is submitted to OGC - Rob Atkinson
  - Compression on server (?) - Rob Atkinson
  - Fix memory leaks (?) - Rob Atkinson
  - Report back [to Rob Atkinson] on NERC Portals experience serving CSML (04/08) - STFC

9 The word(s) of the end

(said by Andrew Woolf)

CSML is not aiming at becoming a data management standard or protocol for ocean/atmosphere but it is now a tool for modelling and harmonising the ocean/atmosphere datasets.

The final objective is to disseminate the concepts (data models) coming out of CSML studies so that the ocean/atmosphere data interoperability is achieved within the OGC standards.

10 Participants :

Name	e-mail	Affiliation
Andrew Woolf	A.Woolf@rl.ac.uk	Environmental e-Science and Spatial Informatics (STFC)
Declan Dunne	D.Dunne@ucc.ie	CMRC
Dominic Lowe	D.Lowe@rl.ac.uk	Environmental e-Science and Spatial Informatics (STFC)
Jean-François Piollé	jfpiolle@ifremer.fr	IFREMER
Keiran Millard	k.millard@hrwallingford.co.uk	HR Wallingford
Mickael Treguer	mickael.treguer@ifremer.fr	IFREMER
Olivier Lauret	olauret@cls.fr	CLS
Stéphane Tarot	Stephane.Tarot@ifremer.fr	IFREMER
Sylvie Pouliquen	Sylvie.Pouliquen@ifremer.fr	IFREMER
Thierry Carval	Thierry.Carval@ifremer.fr	IFREMER
Thomas Loubrieu	thomas.loubrieu@ifremer.fr	IFREMER
Tony Jolibois	tjolibois@cls.fr	CLS
Valerie Harscoat	valerie.harscoat@ifremer.fr	IFREMER